LB 3051
.F5
Copy 1

National Psychology Monographs

Class LB3051

Book .F5

Copyright No. 1913

COPYRIGHT DEPOSIT.

THE MARKING SYSTEM IN THEORY AND PRACTICE

Educational Psychology Monographs

Edited by Guy Montrose Whipple

No. 10

The Marking System in Theory and Practice

By

I. E. FINKELSTEIN, A. M. (Cornell)

(Studies from the Cornell Educational Laboratory, No. 14)

WARWICK & YORK, Inc.

Baltimore, U. S. A.

1913

Monograph

LB3051
.F5

Copyright, 1913
By
WARWICK & YORK, Inc.

$1.00

©Cl.A354358

EDITOR'S PREFACE.

When we consider the practically universal use in all educational institutions of a system of marks, whether numbers or letters, to indicate scholastic attainment of the pupils or students in these institutions, and when we remember how very great stress is laid by teachers and pupils alike upon these marks as real measures or indicators of attainment, we can but be astonished at the blind faith that has been felt in the reliability of the marking system. School administrators have been using with confidence an absolutely uncalibrated instrument. Only within a very few years have serious attempts been made to scrutinize the theory of marking or to test by statistical and experimental procedure the degree of precision that could be expected in its use.

What we need to know is: What are the traits, qualities or capacities that we are actually trying to measure in our marking systems? How are these capacities actually distributed in the body of pupils or students? What method ought we to follow in measuring these capacities? What faults appear in the marking systems that we are now using, and how can these be avoided or minimized?

This monograph (originally prepared as a master's thesis at Cornell University) is a contribution directed toward the answering of these very pertinent questions. In it the author reviews the conclu-

sions reached by previous investigators, sets forth the underlying theories of marking systems, and, finally, demonstrates by a painstaking statistical analysis of the marks given in his own institution what degree of unreliability and what faults of distribution inhere in the ordinary percentile system that is employed in most schools and colleges. These statistical results must not be thought to be peculiar to a particular university, or to universities in general. They will be found, upon examination, to be the pattern to which the marking system of any educational system will tend to conform, and for this reason this study has not a local, but a general significance, and the author's conclusions and recommendations deserve most careful study by all who are concerned in educational administration.

Ithaca, N. Y., April, 1913. G. M. W.

CONTENTS.

CHAPTER I.

Page

INTRODUCTORY 5

CHAPTER II.

THEORETICAL CONSIDERATIONS 9

1. Should marks indicate performance or ability or accomplishment? 9
2. What is the theoretical distribution of the qualities or traits that marks are to indicate? ... 11
3. What is the best method of translating the distribution into a scale of symbols? 16

CHAPTER III.

THE DISTRIBUTION OF MARKS AT CORNELL UNIVERSITY: COMBINED RESULTS FOR NUMEROUS COURSES 21

1. Marks given in 1902 25
2. Marks given in 1903 25
3. Marks given in 1911 26
4. Combined Curve of Marks in 1902, 1903 and 1911 28

CHAPTER IV.

THE DISTRIBUTION OF MARKS AT CORNELL UNIVERSITY: RESULTS FOR INDIVIDUAL COURSES 37

1. Variation produced by change of instructors .. 39
2. Typical distributions of "high markers" 42
3. Typical distributions of "low markers" 49
4. Peculiarities of distribution in other courses ... 60
5. Marking system of the College of Law 72

CHAPTER V.

SUMMARY AND CONCLUSIONS 79

CHAPTER I.

INTRODUCTORY.

The idea of making a careful investigation of the statistical and psychological problems underlying the assignment of grades or marks to students in schools and colleges is of relatively recent date. It is within the last decade that serious attention has been paid to such queries as: What should the mark really represent? Should the mark be based upon ability or performance, or even upon zeal and enthusiasm? What is the best set of symbols to represent ability or achievement? How are the marks given by different teachers or different schools actually distributed? Is it possible, by exhibition of distributions, or by formal instruction in the theory of marking, to increase the fairness and reliability of marks? Do students tend to secure the same standing under different teachers in the same school or to maintain their relative standing when proceeding from class to class or from school to college?

From the studies of J. M. Cattell (1905), W. S. Hall (1906), Max Meyer (1908), W. F. Dearborn (1910), A. G. Smith (1911), A. G. Steele (1911), W. T. Foster (1911) and others[1] who have discussed various phases of these problems has come the demonstration that few teachers stop to consider what the

[1]For exact references see the bibliography, p. 85.

marking system under which they work really implies; that the variability in the marks given for the same subject and to the same pupils by different instructors is so great as frequently to work real injustice to the students; and that the marking system in most common use—the percentage system with 100 for a maximum and 60 or 70 as a "pass mark"—is in all probability not the best system.

If these conclusions be granted, it is evident that the reliability of the marking system in any institution of learning is a matter for investigation. If the teachers in the institution are marking unscientifically, or if they are using a system which can be shown to be inferior in theory and in practice, then these facts should be investigated and a remedy sought. Nor may anyone seek refuge in the assertion that the marks of the students are of little real importance. The evidence is clear that marks constitute a very real and a very strong inducement to work,[1] that they are accepted as real and fairly exact measurements of ability or of performance. Moreover, they not infrequently are determiners of the student's career. They constitute the primary basis for election to honorary societies, for the award of various academic honors, for advancement from class to class, for graduation, and may even determine in some measure the student's career after leaving the institution in which they have been assigned.

As Meyer (9, p. 661) remarks, "If different grades were simply means of giving some students notoriety above others, the question would be immaterial, for a gentleman does not seek notoriety. But the

[1]Colvin (2) shows how the marking system can be used as an incentive if it is well organized and rational.

grade has in more than one sense a cash value, and if there is no uniformity of grading in an institution, this means directly that values are stolen from some and undeservedly presented to others.

"The result is that, among the members of the faculty as well as among the students, men look at each other with suspicion. That this attitude is detrimental to the feeling of unity, to the development of a college spirit, is clear to even the most superficial observer. Whatever contributes to a greater uniformity of grading contributes directly towards more peace, a better mutual understanding, a greater community of purposes among all the members of the institution."

The purpose of the present study is primarily to examine the distribution of marks as found in various colleges and classes in Cornell University. But consideration is first given to the relative merits of different marking systems, and to the theoretical considerations which underlie any scientifically organized system.

CHAPTER II.

THEORETICAL CONSIDERATIONS.

Three theoretical problems deserve consideration before we set forth the data obtained from our investigation of the actual distribution of the marks of Cornell University.

The first problem is: Should marks indicate performance or ability or accomplishment? The second problem is: What is the theoretical distribution of the quality or traits that the marks are to indicate? The third problem is: What is the best method of translating the distribution into a scale of marks?

1. *Should marks indicate performance or ability or accomplishment?*

In certain cases, where the examiner has before him merely the results of the examinee's efforts to answer a given set of questions or to solve a given group of problems, as, for instance, in the examinations submitted in the Civil Service, it is evident that marks must be based upon performance, *i. e.*, upon work actually done in the examination, without regard to native ability or zeal or previous evidence of acquaintance with the subject-matter of the examination.

But in actual school or college work, the teacher has more to guide him than performance in examination alone. He is able, as a rule, to form some idea

of the pupil's native ability. He is able, furthermore, to take into account evidence afforded in other ways than by the performance, of the pupil's real knowledge and acquaintance with the subject-matter of the course. Thus, he may be convinced that a certain pupil fails to do himself justice in his classroom and examination performance. He may then decide to raise his mark in such a way as to indicate more fairly his accomplishment (as distinguished from mere performance).

The issue then takes the form: Shall the pupil be marked according to his ability or according to his accomplishment?

By ability we mean native endowment, intellectual capacity. Accomplishment is certainly very largely determined by ability, but it is also determined by adequacy of previous preparation, and perhaps still more by zeal and effort. College teachers will readily recognize the type of student whose native ability is handicapped by poor training in the preparatory school, and all teachers recognize the type of student whose ability is not reflected in his accomplishment for lack of earnest application.

To most teachers it seems axiomatic that marks should indicate accomplishment, and not ability alone. If a capable student shirks his work in physics, he must suffer the penalty of a low mark. If a dull student passes an examination successfully by dint of strenuous application, he is entitled to the credit of his accomplishment.

But the opposite position has been defended by some writers. Thus, A. G. Smith (11, p. 384) argues that "College men have greater mental accomplishment than the average man, but, measured from the

standpoint of mental ability or capacity, they are not a group so highly specialized as is commonly believed. This is especially true in America, where the colleges are filled with students drawn from every walk of life." He asserts that college grades, when properly given, should represent ability rather than accomplishment.

2. *What is the theoretical distribution of the qualities that marks are to indicate?*

Let us, for the moment, keep both of these positions in mind. What, now, should be the distribution of marks of a class, first, when the marks indicate ability; secondly, when the marks indicate accomplishment?

Native ability, from all the evidence at our command, behaves like any other biological trait. It follows that in any population its distribution is that known as the curve of error, the probability curve, or Gauss's curve. This curve, *Fig. A*, is a bell-

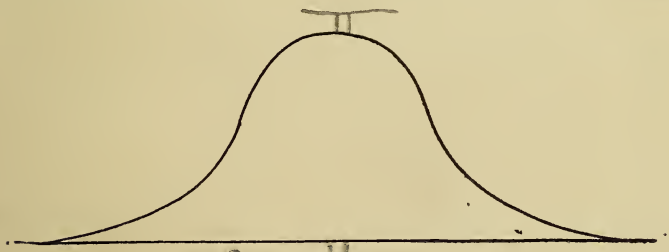


FIGURE A.

shaped symmetrical curve, with a mode at the median point, with deviations of equal magnitude and frequency above and below the median, and with a progressive diminution in frequency of occurrence as the deviation increases in magnitude.

The question now arises whether this curve represents truly the distribution of ability in high-school and in college students (with whom we are more particularly concerned). The chief factor which might invalidate the application of this curve to a class of any rank above the primary grades is the factor of selective elimination. Evidently, the idiot, the imbecile and the moron are eliminated before the high school is reached—the idiot at the outset, the imbecile early in the grades, the moron perhaps in the grammar grades or earlier. Undoubtedly, there occurs at points in the public school system above the grammar grades a still further elimination of dull pupils from the lower end of the curve of distribution. Those who cannot win promotion in the high school cannot reach the college. If, then, the lower end, at any rate, of the curve of distribution of native ability for the entire population be thus cut off by the machinery of the public schools, it might appear that the distribution of native ability in the college would resemble the curve shown in *Fig. B,* which is obviously skewed to the left.

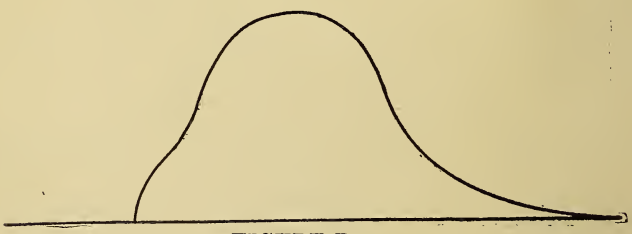


FIGURE B.

Not every one, however, is agreed that elimination of this kind does take place. It would appear that the elimination is not so severe, at any rate, as is

often supposed. Judd (7, p. 469), for instance, says: "A study of the large high schools of the city of Chicago in its relation to the University of Chicago shows that students go to college from every level of scholarship above the passing mark." Similarly, Meyer (9, p. 666) writes: "College teachers usually assert that the curve of distribution is not the normal curve, but a skewed curve. . . . They are usually ready to explain this by referring to the elimination of poor scholars in the high schools and lower schools. I have considerable doubts as to this elimination. Is the work done in a high school really so much like that done in college that there is a large previous elimination of poor college students?"

However, the situation, in our opinion, is not quite so simple as this. As the pupil passes from kindergarten to university the standard of ability presupposed and exacted for successful work gradually rises. So, also, does the ability of the pupil, and this, merely from maturity, quite independently of his instruction. Along this path of progressive development of ability arrests occur. The feeble-minded are instances of such an arrest in the lower stages of mental development. Modern psychology teaches us that "retardations occur continuously on up the years of growth to maturity" (Huey, 6, p. 43). The mean ability of a group of children of a given age, therefore, advances with increasing age. The distribution of capacity of a thousand children at their tenth year may, therefore, resemble that of the same children at their fifth year, save that the entire group tends to move forward to a higher stage.

When the remnant of this same group arrives at college, its arrested members have dropped out by

elimination, but the poorest of the surviving members represent a minimal ability just sufficient to cross the deadline of university entrance. The distribution, then, of this group of college freshmen might still be that of the normal probability curve, save, of course, that the standard of the whole distribution now differs from that of the same group when pupils in the kindergarten, just as the age-norms in the Binet-Simon scale advance progressively in difficulty. If these contentions be accepted, the elimination of the academically unfit by the mill of the public school system does not produce a distribution like that of *Fig. B,* but leaves us, after all, the form of distribution shown in *Fig. A.*

Hence, if college marks should indicate ability, Meyer, Smith, Dearborn, Cattell and others would undoubtedly be right when they aver that the probability curve should represent the type or pattern of the normal curve for the distribution of university marks, and it would then become a relatively simple matter to lay down rules for the guidance of instructors and for the standardization of the marking system for any institution, as has been recently done in the University of Missouri.

But now let us return to the second possibility, viz., that marks should indicate, not ability, but accomplishment. Is accomplishment distributed like ability? We have argued that accomplishment is the result of ability, plus previous preparation and zeal. When like zeal or effort is exhibited by two pupils, the one dull and the other brilliant, it seems probable that the gifted pupil reaps a *proportionately* larger result. It is as if effort were multiplied into ability rather than added to it. It is difficult to

prove this assertion, but if it be admitted, the conclusion is evident that, if all the students put forth the same amount of effort, the curve of accomplishment would be skewed to the right (*Fig. C.*). This

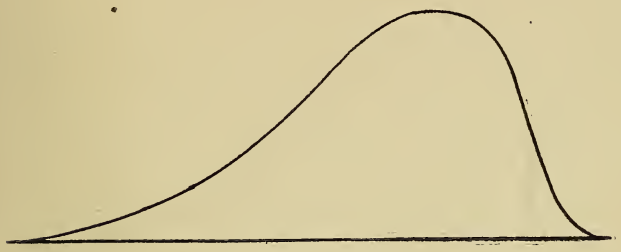


FIGURE C.

skew would appear in the absence of any marking system. But when a marking system is arranged to measure accomplishment, another factor undoubtedly enters to skew the curve further to the right. This factor is the incentive offered to certain groups of students by the critical points of the system. Thus, at Cornell University, where a mark of 60 or above is necessary to pass, there evidently exists a type of student of inferior or average ability who aims for this mark. The possibility of exemption from final examination by attaining a mark of 85 in preliminary examinations offers also a powerful incentive to students of average ability to push their accomplishment to the extent of their ability. Again, an average student, by persistent effort, may hope to obtain an average of 80 or over, and thus be eligible for consideration for Phi Beta Kappa and other special honors. On the other hand, the brilliant student, who probably profits most by effort, cannot at-

tain a mark above 100, and it is not difficult for him to obtain a mark of 85 or over. The net result, then, of the imposition of a marking system is to crowd the grades of accomplishment forward to the right, and thus, again, to tend to form a skewed, and not a bell-shaped curve.

Finally, the actual distribution of over 20,000 marks in Cornell University is a curve skewed to the right, as will be shown below (p. 28). If such a curve, compounded of many classes and many examiners, may be thought to represent with great reliability the *consensus gentium* of the faculty of the University with regard to the distribution of accomplishment, it affords us one more reason for believing that the distribution of accomplishment is skewed to the right. The pattern or theoretically ideal curve of high-school and college marks is, therefore, not the probability curve, but the skewed curve with the mode to the right of the middle of the abscissa.

3. *What is the best method of translating the distribution into a scale of symbols?*

Given now a pattern or an ideal distribution of accomplishment, the further question remains: What is the best method of dividing this distribution into groups for translating accomplishment into a symbol or mark? Theoretically, there are numerous ways of making such a division, of translating standing into a scale of marks. In actuality serious consideration has been given to a few forms of scale only, viz., the division into two, into three, into four, into five and into one hundred groups.

(*a*) *The two-division system.* Simplest of all devices is that which divides all students into two

groups—"passed" and "not passed." Such a division is in operation at Cornell University, and at many other institutions in the reporting of the work of students in the Graduate School. While more definite marks may be assigned at the discretion of the instructor, it is customary to report the work of the students simply as satisfactory or unsatisfactory. There are some members of the faculty who would be glad to see this system extended to undergraduates, but the majority of college teachers, and of students as well, prefer to use a more precise scale, despite the greater labor entailed in the grading of work and in the keeping of records. It is felt that the student should receive a more exact notion of his accomplishment, and that for many extraneous purposes—selection of members of advanced classes, distribution of various awards, etc.—a finer scale is necessary.

(*b*) *The three-division system.* E. B. Sargent (10, p. 64), who has advocated the placing of a pattern curve of distribution on all record sheets to guide the marking of examiners, argues for a three-division system in which the groups are labeled: inferior, mediocre and superior. The fundamental merit of this system is the psychologically correct distinction of a large group of students of average accomplishment, midway between two smaller groups whose work falls short of, or exceeds, this average accomplishment. The defects of this system lie, first, in the difficulty of distinguishing failure from mere inferiority which is still entitled to a pass, and, secondly, as A. G. Smith (11, p. 390) has pointed out, "that it furnishes no distinguishing mark of excellence unless the middle group is made

so large as to be open to all the criticism that can be urged against a 'pass' and 'not pass' method of grading."

(*c*) *The four-division system.* No one has very seriously defended a four-division system. There is no psychological nor statistical justification for it.

(*d*) *The five-division system.* The best possible division of the marking scale for any small number of groups is the five-member division. This plan is based upon the orientation of all students around a central group whose accomplishment is construed to be average or medium. The theory of the distribution of ability, and hence in large measure of accomplishment, teaches us that mediocrity is the commonest condition. The largest single homogeneous type of student is the average student. Above and below the average lie groups of smaller size containing superior and inferior students—superior and inferior with reference to the average group. The five-division system improves upon the three-division system in that it further differentiates the outlying groups; those superior to the average are subdivided into two groups, the superior and the excellent or exceptionally good; those inferior to the average are subdivided into inferiors and failures.[1]

The theory of the application of the five-division system to the actual grading of students assumes that the actual distribution of marks should conform

[1]For purposes of administration it may be thought desirable further to differentiate between 'conditioned' and 'absolutely failed'—following the custom of many institutions of permitting the former to try for a 'pass' by taking a 'make-up' examination, but compelling the latter to take the course again in its entirety in order to secure credit. In practice the scale would then contain six symbols, but it would, nevertheless, be in theory a five-division system.

fairly closely to a theoretically predetermined distribution. The question as to what this predetermined distribution should be in the case of a skewed curve will be treated later (pp. 30-33). We may set forth here, however, the plans proposed by certain psychologists for the translation into marks of the bell-shaped curve of distribution. If we assign the symbols *A, B, C, D* and *E* to excellent, superior, average, inferior and failure, respectively, the divisions recommended by Professors Meyer, Dearborn and Cattell for each hundred students are as follows:

	A.	B.	C.	D.	E.
Meyer	3	22	50	22	3
Dearborn	2	23	50	23	2
Cattell	10	20	40	20	10

In the opinion of these writers, then, from 40 to 50 per cent. should be marked average, from 20 to 23 per cent. should be superior, and inferior to the average, and from 2 to 10 per cent. should receive the highest mark and the like number should fail.

(*e*) *The percentile system.* It is not very difficult to grade students on the five-division system. Is it possible to say as much of the system which many institutions follow, according to which marks are based on a scale of 100 points? Theoretically, this scale implies that distinctions of a fineness of one-hundredth may be made, and in practice such distinctions are constantly attempted. But what is the difference, if any, between a mark of 75 and one of 76? What, for that matter, does 75 mean? Has the student accomplished 75 per cent. of some ideal accomplishment? It is a commonplace of statistics that a scale whose units are not defined or whose units are not identical throughout is no scale at all.

The fact that different instructors place a different interpretation upon the symbols of the percentile system is evidence enough that it is not the scientific measuring rod that it pretends to be. The very fact that its divisions are so minute is doubly insidious; it promises precision, but it cannot afford it.[1] In short, the 100-division scale has no psychological justification. On the other hand, the five-division system, which is evidently based on a different plan altogether, is simple to use, and the results of each instructor are easily checked at any time.[2]

[1]Since this was written, Professor Starch of the University of Wisconsin has reported the results of a study upon the "Reliability and Distribution of Grades" (*Psychol. Bulletin*, 10, Feb. 15, 1913, p. 74), which shows that the marks assigned by more than one hundred teachers to two papers in English and one in geometry have a probable error of from 4.0 to 7.5 points on a percentile scale. Starch concludes: "The steps on a scale should be at least twice the size of the mean variation or probable error of the measurements in order to be distinguishable steps. Hence the steps on a marking scale should be at least two times 4.2, or approximately 8 points. And hence on a scale of passing grades of 70 to 100 only four steps can be used with any degree of objective reliability." This statistical conclusion, then, confirms very prettily our argument for a five-division scale—four marks above the passing limit and one mark below it for failures.

[2]See, for example, the very interesting history of the establishment and operation of a five-division system at the University of Missouri, as narrated by Meyer (8, 9).

CHAPTER III.

THE DISTRIBUTION OF MARKS AT CORNELL UNIVERSITY: COMBINED RESULTS FOR NUMEROUS COURSES.

The purpose of this chapter is to show the actual distribution of marks at Cornell University when a large number of different courses are combined. The effect of combining the marks of several thousand students in numerous courses is, naturally, to eliminate or to cancel *chance* variable errors in the marking. The resultant curve thus compounded of the marks given in varied classes is as true a picture as can be obtained of the actual distribution of accomplishment of students as judged by the instructing staff. Whether the distribution may not be affected by certain *constant* errors is a matter to be discussed a little later.

The material at our disposal for securing these combined results comprises three sets of data. The first set represents the marks given in the College of Arts and Sciences during the first term of 1902-03 to 5396 students in 66 courses. The subjects represented are Latin, German, French, English, Philosophy, Psychology, Education, History and Political Science, Mathematics, Physics, Chemistry, Botany, Invertebrate Zoology, Physiology and Geology. The second set of data represents the marks given in the same College in the following academic

year. It includes the same courses and 7522 students.[1] The third set of data represents the marks collected by the writer and shown in detail in Chapter IV. It includes 7430 marks in 31 courses in different colleges of the University (Arts and Sciences, Agriculture, Mechanical Engineering and Civil Engineering), and also 711 marks in three courses in the College of Law. These three sets of data will be referred to for convenience as the 1902 marks, the 1903 marks and the 1911 marks.[2] We shall present the distribution for these three sets of data in the order given, then combine them into a single distribution and discuss the form of this final compounded curve.

TABLE I.

COLLECTIVE DISTRIBUTIONS.

Showing the Per Cent. of Students in the Several Groups.

Year.	No. of Marks.	0-39.	40-44.	45-49.	50-54.	55-59.	60-64.
1902.....	5,396	1.6	1.7	1.6	3.4	2.3	11.8
1903.....	7,522	1.3	1.2	2.1	3.5	1.8	10.2
1911.....	7,430	0.6	0.4	1.1	2.7	2.2	12.8
All three.....	20,348	1.2	1.1	1.6	3.2	2.1	11.6
All three. (Revised)....	20,348	1.2	1.1	1.6	3.2	5.6	8.1

Year.	65-69.	70-74.	75-79.	80-84.	85-89.	90-95.	95-100.	Exempt.
1902.....	10.2	12.9	15.8	15.6	11.5	8.9	2.7	23.1
1903.....	10.0	13.6	16.0	16.2	11.6	8.7	3.8	24.1
1911.....	12.5	15.4	16.4	12.7	14.0	7.2	2.0	23.2
All three.....	10.9	13.9	16.1	14.8	12.4	8.3	2.8	23.5
All three. (Revised)	10.9	13.9	16.1	14.8	12.4	8.3	2.8	23.5

[1]My thanks are due to Prof. W. F. Willcox of Cornell University for the use of these two sets of data which were compiled under his direction several years ago.

[2]As explained in Chapter IV, this third set of data contains marks extending backward from June, 1911, for a length of time necessary to secure at least 200 marks for each course. In most cases the period represented falls between 1910 and 1911.

To make the charts intelligible a word of explanation is needed. The percentile system is used at Cornell—save in the College of Law, of which we shall speak more definitely below. The 'pass mark' is placed at 60. A mark between 40 and 59, inclusive, is known as a 'condition'—the student is entitled within one year to try a 'make-up' examination in order to gain credit for the course, provided he reaches 60 or above in this examination.[1] A mark below 40 represents complete failure; the student must take the course again in its entirety, and must then pass at 60 or over in order to gain credit for his work. In certain courses students may, at the option of the instructor, be exempted from the final written examination. The mark which must be attained to secure exemption varies in different colleges, save that in the College of Arts and Sciences exemption, if given at all, must be based upon a mark of 85 or over. In the College of Mechanical Engineering there are no special or formal final examinations; the mark is based upon the work of the students during the term.

All the charts of distribution are plotted with the scale of marks as the abscissas, and in units of 5 (or 6) points each, with the highest marks at the right. Thus, the division at the extreme right represents the six marks 95, 96, 97, 98, 99 and 100. The next division represents the five marks from 90 to 94, inclusive, and all subsequent divisions on toward the left end of the abscissa each represent five marks, save that all marks including 39 and below are in

[1]Through oversight, it wos not noticed until too late to make the correction that the mark of 40 is counted as complete failure, not as conditioned. This error does not affect the conclusions in any way — *Editor.*

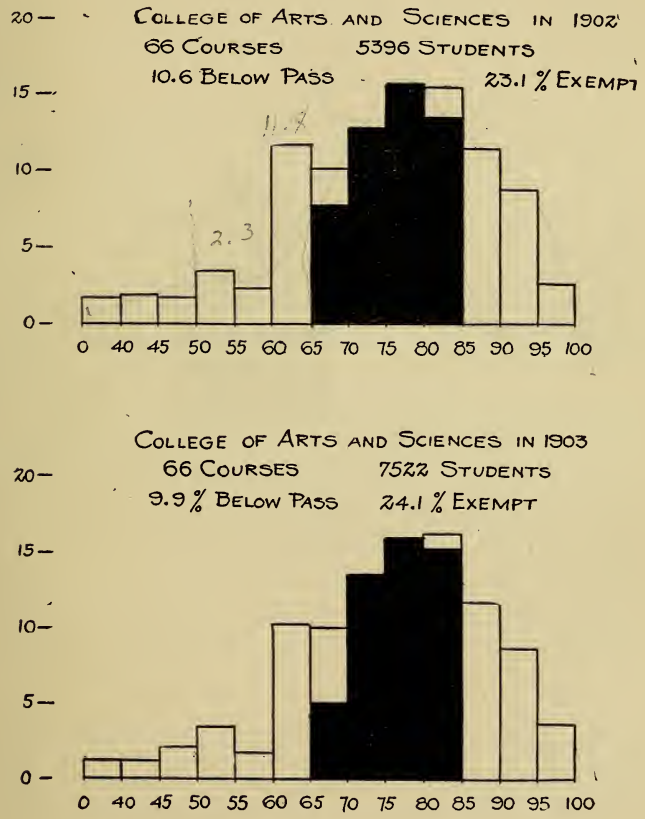
the single division at the extreme left. The ordinates in all charts represent the per cent. of students receiving the marks in the various divisions of the scale.

In all the charts, furthermore, two additional features of the distribution are shown. First, the *percentage* of students who would be exempt from the final examination, on a basis of 85 or over (assuming exemption were permitted in all classes), is shown in the accompanying legends. Secondly, the range of marks received by the middle 50 per cent. of the students is shown on each graph by the size and position of the solid black area. This range is found by counting off 25 per cent. of the marks from the upper, and 25 per cent. of the marks from the lower end of the total distribution found in each case.[1]

Since, by theory, the middle 50 per cent. of any group of students must be neither brilliant nor dull, but simply straightforward, average students, the distribution of this group is of special interest. Speaking generally, we should not expect the marks obtained by this group to spread over a large range, since they represent a homogeneous group in point

[1]To avoid possible misunderstanding, a word should be said concerning the boundaries of the black area on the charts. The upper and lower limits of the middle half of the students need not necessarily, of course, coincide with a division point upon the abscissa. Thus, in Chart I, the upper limit of the middle group lies in the range 80-84, but there are also some students in the upper 25 per cent. whose marks fall within the same range. To be specific, since 23.1 per cent. lie in the range 85-100, 1.9 per cent. of the students superior to the average lie in the range 80-84 per cent. Accordingly, the vertical column erected as an ordinate over the range 80-84 is blackened to within 1.9 units of its tip only, and not fully to the tip. Similarly, a mark lying between 65 and 69 is for the most part obtained by students belonging to the middle or average group, but there are a few students inferior to the average (on our theoretical definition of average) that also obtain marks within this range.

of ability and accomplishment. Again, we should not expect an average student to attain exemption from final examination, nor, on the other hand, to run any grave risk of being 'conditioned.'



CHARTS I AND 2. Combined Distribution of Marks in 66 Courses in 1902 and 1903.

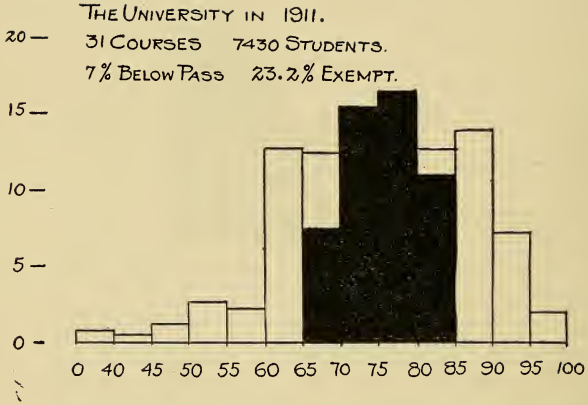


CHART 3. DISTRIBUTION OF MARKS IN 31 COURSES IN 1911.

The distribution of the three sets of data used for combined curves is shown in Table I, in tabular form, and in Charts 1, 2 and 3, in graphic form. Chart 4 is the combination of Charts 1, 2 and 3.

It will be noted that the distribution in the three sets of data is closely similar. They all take the form of a curve skewed to the right, and with an evident disturbance in the region of 60, the 'pass' mark.

The combination shown in Chart 4 is worthy of special mention. It includes in all 20,348 marks, enough surely to entitle it to the designation as the true curve[1] of the distribution of accomplishment, as indicated by the marks of Cornell University. The form of the curve is again the skew to the right, not the probability curve which, according to Meyer, Cattell, Dearborn and others, is the *theoretically* correct distribution.

In Chart 4, just as in all the three curves from which it is compounded, there appears a disturbance at 60. There are, apparently, too few marks between 55 and 59, too many marks between 60 and 64. Two factors may contribute to this result. First, as suggested already, it is probable that a certain number of students make a special effort to just secure a pass mark. Secondly, many instructors are loath to

---

[1]A. G. Steele (12, p. 525), from an experiment conducted during a summer session at Miami University, concludes that "the average judgment of several competent judges gives approximately the true grade, and the one which will be more nearly rightly interpreted by the greatest number of judges."

Similarly, in the experimental study of the psychology of testimony, it has been found that the combined result of the evidence of several witnesses affords a fairly reliable indication of the true state of affairs.

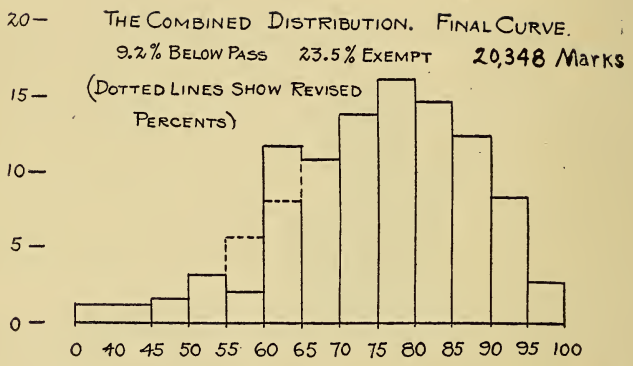

THE COMBINED DISTRIBUTION. FINAL CURVE.
9.2% BELOW PASS 23.5% EXEMPT 20,348 Marks
(DOTTED LINES SHOW REVISED PERCENTS)
20
15
10
5
0
0 40 45 50 55 60 65 70 75 80 85 90 95 100

CHART 4. COMBINED DISTRIBUTION OF THE 20,348 MARKS REPRESENTED IN CHARTS 1, 2 AND 3.

condition a student by a narrow margin. If the work of the term justifies a mark between 55 and 59, the instructor frequently raises the mark, arbitrarily, to 60. He may feel a measure of insecurity in his judgment, and he may dislike to suggest to the student the possibility of any argument about the 'condition' which a mark in the region of 58 would entail. The second of these two factors is undoubtedly the more potent in distorting the distribution in the region of 60.[1] The disturbance in the 60 region has been eliminated by the method of differences and the resulting correction is shown by the dotted lines of Chart 4.

In Chart 1 there is 2.3 per cent. in the 55-59 group, and 11.8 per cent. in the 60-64 group. Let 11.8 be equal x. Then the 55-59 group is equal to 14.1-x. Taking two figures on each side of these numbers, we find x in the fifth difference.

1.6	3.4	14.1—x	x	10.2	12.9	
1.8	10.7—x	$2x$—14.1	10.2—x	2.7		
8.9—x	$3x$—24.8	24.3—$3x$	x—7.5			
$4x$—33.7	49.1—$6x$	$4x$—31.8				
82.8—$10x$	$10x$—80.9					
$20x$—163.7						

x is, therefore, equal to 8.1.

The 60-64 group should then be 8.1, and the 55-59 group should be 6.0. In the same way we smooth the graphs of Charts 2 and 3. For Chart 2 we find that the 60-64 group should be 7.1 instead of 10.2, and the 55-59 group 4.9 instead of 1.8. For Chart 3 the 60-64 group is changed from 12.8 to 9.2, and the

[1]There are some instructors who deliberately raise the mark of the doubtful cases, not merely to 60, but to 65 or even to 70, with the idea that if a student is to be passed, he "might as well be passed handsomely," as one of them puts it.

55-59 group is changed from 2.2 to 5.8. In Table I the data of Chart 4 appear in tabular form both with the original and the revised per cents.

A word should now be said about the average mark and the translation of the curve of Chart 4 into a five-division marking system distributed around the average group. Here we recur to the problem broached in Chapter II (p. 19), but now with the advantage of an actual curve of distribution as the basis of discussion.

The weighted arithmetical mean for these several compounded curves is as follows: 1902 marks, 73.0; 1903 marks, 74.12, and 1911 marks, 75.23. We may be justified, then, in taking 75 as the net average mark of the students here represented for the purposes of fitting the curve of actual distribution to the curve of theoretical distribution. Now, if we assumed, like Cattell, Meyer and others, that the symmetrical probability curve were the correct theoretical distribution, then, if 75 were the median or mode, there ought to be equal divisions above and below this mode. But the deviations above 75 cannot exceed 25 points; hence the lowest possible mark, or complete failure, would have to be indicated by the mark of 50.

It is perfectly evident that this conclusion leads to an absurdity. Complete ignorance of a subject should be represented by a mark of 0, according to the implication of the percentile scale. It is true that few marks are given below 50; nevertheless, such marks are given, and, furthermore, a distinction is made, administratively, between marks below 40 and marks between 40 and 59.[1] If a translation of the normal probability curve to the percentile scale

[1]See note, p. 23.

is to be made, then the meaning of the percentile scale must apparently be radically altered.

If, however, we assume that the correct theoretical distribution, for the reasons advanced in Chapter II, is a curve skewed to the right, then the translation of this curve to the actual distribution shown in Chart 4 is feasible. Let us start with the criterion for exemption, the mark of 85. On all accounts, an accomplishment sufficient to justify exemption from final examinations should be of a grade superior to the average accomplishment. Exemption should be a privilege reserved for superior work. From Chart 4 we find that 23.5 per cent. of the students of the University reach this degree of accomplishment. This number corresponds, then, very closely to the number that Dearborn and Meyer assign as the number that are theoretically superior to the average group.

Secondly, our distribution suggests a differentiation of the superior group at the mark of 95. A mark of 95 or over is gained by 2.8 per cent.; a mark of 85 to 94, inclusive, is gained by 20.7 per cent. of the Cornell students. This division into excellent and superior students also corresponds very closely to the distribution assigned from theoretical grounds by the writers just mentioned.

Thirdly, in the system used at Cornell the mark of 60 serves as another crucial point. A mark below 60, that is, a 'condition' or a 'failure,' is given in actual practice to 9.2 per cent. of the students. Here we are unable to arrange a distribution that corresponds to the theoretical requirements of the probability curve according to Meyer and Dearborn, but the distribution does closely correspond to Cattell's

assignment of 10 per cent. to the group of poorest students. But we have already seen that the marks between 55 and 59 are affected by the operation of a constant error. Our correction of this error (see p. 29 and Chart 4) brings the number properly assignable to this lowest group to 12.7 per cent. Of these marks a subdivision may be made which will give 1.2 per cent. to 'complete failures' and 11.5 per cent. to 'conditions.' In fine, then, it does not appear that the poorest group should be equal in size to the best group; rather that the number of students conditioned and failed should properly exceed the number that receive the mark of highest excellence.

Fourthly, we have left a group of 63.8 per cent. of inferior and average students whose work entitles them to pass, but does not entitle them to exemption from final examination. The problem is to determine the dividing line between inferior and average accomplishment. Theoretical considerations compel us to assign more students to the average than to the inferior group. Inspection of Table I and Chart 4 shows that a division at 70 yields satisfactory results. The inferior group, now receiving marks between 60 and 69, inclusive, comprises 19.0 per cent. of the students, and is thus practically equal with the superior group previously set aside. The average group, now receiving marks between 70 and 84, contains 44.8 per cent. of the students. As noted, Meyer and Dearborn would place 50 per cent. and Cattell 40 per cent. in this group.

To eliminate the decimals, we may lay down as

the pattern or ideal distribution of marks the following schema:

Group.	*Poorest.*	*Inferior.*	*Average.*	*Superior.*	*Excellent.*
Per cent.....	12	19	45	21	3
Range in percentile scale }	0-59	60-69	70-84	85-94	95-100

In our judgment, it would be in every respect desirable for Cornell University, and any other institution of like character, and probably also for the secondary schools as well, to adopt a five-division system of marking with the express provision that, in the long run, the marks given by any instructor must not deviate widely from the distribution just indicated.

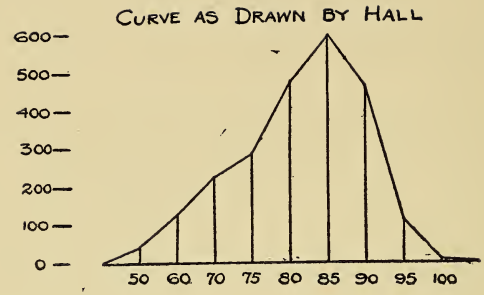
An example of a translation of a distribution of marks into a theoretical curve is afforded by the work of Dr. W. S. Hall of the Northwestern University Medical College. Hall's data are quoted by subsequent writers as evidence of a distribution in the binomial curve. We have been struck, however, with the fact that Hall's distribution is really a skewed curve, like our own, and that it is only by very arbitrary treatment that he has succeeded in exhibiting a semblance to the probability curve. We believe his results confirm our own, and for that reason will consider them briefly at this point.

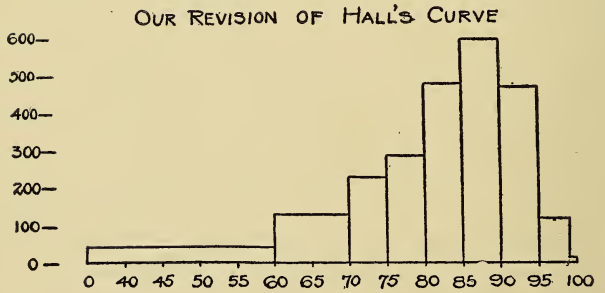
Hall's data are derived from the marks of 2334 medical students, who were marked by a system of nine letters. Hall then translates these nine symbols into a percentile scale (see Chart 5). With this curve before him, he says: "That the curve derived from the rating of the 2334 students is really a binomial curve no fair-minded judge would for a moment question or doubt. We have, therefore, demonstrated beyond cavil that examination data is [*sic*] biologic data and obeys the laws of distribution of biologic data."

In the next breath, however, he adds: "Certain important divergencies from strict coincidence remain yet to be explained. Why does the apex of the curve stand to the right of the symmetrical binomial curve, *i. e.*, why is the curve of my ratings unsymmetrical? The answer is to be sought in two directions:

"1. Either the examiner was too generous and habitually rated his students above their equitable deserts, or

"2. The students were (in a sufficient number of individual cases to influence the totals) guilty of raising their rating above what it should be by nature through dishonest means or extraneous aids in





CHARTS 5 AND 6. Distribution of 2334 Marks Reported by Dr. W. S. Hall.

quizzes, examinations and the preparation of note-books."

As a matter of fact, Hall's curve is not a binomial curve, but a curve skewed to the right. This will be seen clearly enough by reference to our Chart 6, in which we have redrawn the distribution of his data upon a correct scale.[1] It is evident that Hall condensed the left end of his curve unfairly by changing the units of his abscissa at this point, for he has made the ranges 50-59 and 60-69 equal in extent to the other ranges of *five* points each. It is hard to understand how those who quote Hall with approval could have been misled by this feat of statistical juggling.

Finally, the pass mark in Hall's schema is placed at 70. If this point were placed at 60, as is done at Cornell, his distribution would evidently vary still more in the direction of our own pattern curve.

[1]In Charts 5 and 6, ordinates represent absolute numbers, as the appended scale clearly shows.

CHAPTER IV.

THE DISTRIBUTION OF MARKS AT CORNELL UNIVERSITY: RESULTS FOR INDIVIDUAL COURSES.

The purpose in the present chapter is to exhibit the distribution of marks given in 1911, course by course.

The material here presented was gathered by the writer directly from the records of the Cornell University Registrar. The plan was to begin at the records submitted in June, 1911, and to go back in each course until at least 200 separate marks had been secured.[1] Most of the classes were large, so that, save for a few instances, the requisite number of marks was obtained within the two academic years 1909-1911. The charts presented in this chapter are based upon the same abscissas as in Chapter III, and the actual number of marks has been reduced to per cents, so that direct comparison may be made between the different charts. The frequency of exemption is again based on a mark of 85, and the solid black area, as before, shows the distribution of the middle 50 per cent. of the students. The total number of marks is 8141 (including here the College of Law). The range of courses covered is as follows: College of Civil Engineering, one course;

[1] In five instances, however, there are fewer marks than 200; in one of these only 112 marks could be secured; in the other four the numbers lie between 190 and 200.

College of Agriculture, one course; College of Mechanical Engineering, three courses; College of Law, three courses, and College of Arts and Sciences, 26 courses. In the last named group the following subjects are included: Geology, Biology, English, History and Political Science, German, French, Mathematics, Physics, Chemistry, Botany, Education, Psychology and Philosophy.

TABLE II.

EFFECT OF PERSONAL EQUATION AND DISTRIBUTIONS OF HIGH MARKERS.

Showing the Per Cent. of Students in the Several Groups.

Course.	No. of Marks.	0-39.	40-44.	45-49.	50-54.	55-59.	60-64.	65-69.	70-74.
A1.....	263	.4	.4	2.3	4.5	1.9	13.7	11.8	16.7
A2.....	257	...	...	1.2	.8	1.5	12.1	12.8	10.5
B.....	208	...	...	...	...	...	3.8	2.3	2.7
C.....	192	.5	...	...	1.0	...	5.2	3.7	6.3
D.....	293	...	...	...	...	...	...	.7	1.3
E.....	216	...	...	...	...	...	...	.9	2.7

Course.	75-79.	80-84.	85-89.	90-94.	95-100.	Exempt.
A1.....	15.6	20.2	6.4	5.7	.4	12.5
A2.....	13.6	9.8	33.9	3.8	...	37.7
B.....	10.4	13.0	23.1	26.4	18.3	67.8
C.....	5.7	6.8	42.7	16.1	12.0	70.8
D.....	4.4	17.4	24.9	38.6	12.7	76.2
E.....	8.8	9.6	29.2	41.2	7.4	78.0

In carrying out the general purpose of displaying the variability (and hence the presumptive unreliability) of the distribution of marks in specific courses, we shall arrange the material in five parts. The first part will show the variation produced by the change of instructors in a given course; the

second will present typical distributions of "high markers;" the third, typical distributions of "low markers;" the fourth will show numerous peculiarities of distribution in other curves, while the fifth will deal with the special problem of the marking system used by the College of Law.

1. Variation produced by change of instructors.

A certain course in the College of Arts and Sciences continues throughout the year; the work of the first term is in charge of one professor, while the work of the second term is in charge of a different professor. The students, with very few exceptions, are the same in both terms. Chart 7 shows the distribution of marks in the first, Chart 8 in the second term of this course, when the marks of several years are combined. The corresponding numerical data are shown in the first two distributions (Course A1 and Course A2) of Table II. The difference in the two distributions is at once apparent. In the first term 12.5 per cent., in the second term 37.7 per cent. of this class is exempt from final examination. In the first term 0.4 per cent of the students secure a mark above 95, in the second term no student reaches this mark. In the first term 9.5 per cent., in the second term 3.5 per cent. of the students are conditioned or failed. In the first term no student of average ability receives a mark above 84, while in the second term numerous students of average ability receive marks between 85 and 89. In fact, the number of students in this particular range is absurdly large.

This decided difference in distribution of the accomplishment of the same students in the same subject in the two halves of the course is evidently

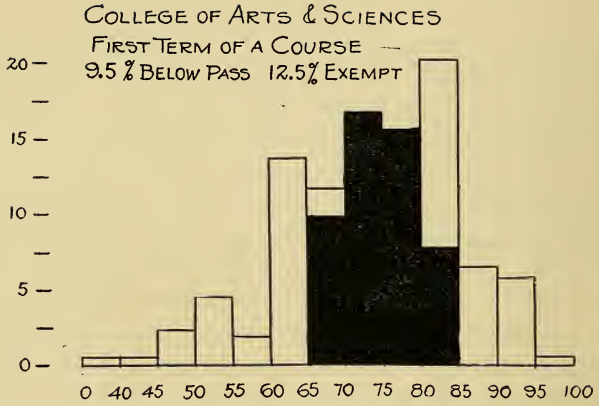


CHART 7. DISTRIBUTION IN COURSE A1.

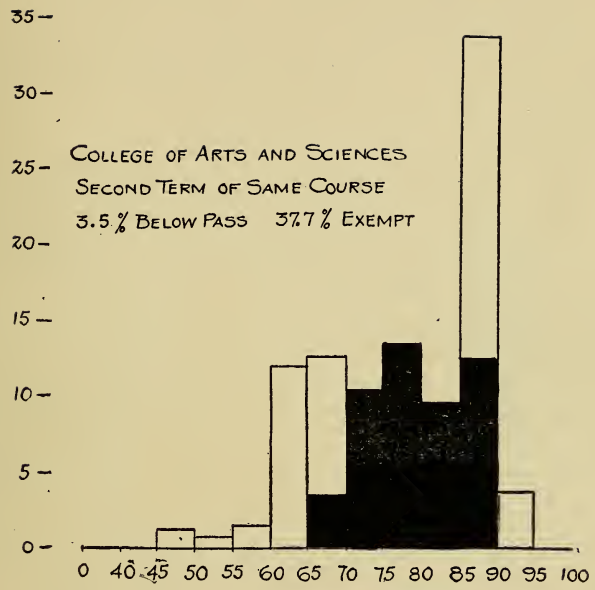


CHART 8. Distribution in Course A2, Second Term of A1.

almost entirely due to the different standards of marking held by the two professors.

2. Typical distributions of 'high markers.'

Charts 9 to 12 exhibit the distributions given by four typical 'high markers.' The numerical data for these distributions are shown in the last four lines of Table II. The reason for these unexpected deviations from the pattern distribution (Chart 4) is apparently different for each case.

Chart 9 represents the distribution of a course in the College of Arts and Sciences in which there are held weekly examinations, but no final examination to cover the work of the entire term. It cannot be said positively that this fact explains the high marks, but it is at least a possible explanation. The students in the course must organize the material of the four lecture periods in preparation for the written exercise which is held in the fifth period of each week. But they need not organize the material of the entire course in preparation for a long formal final examination. Whatever be the explanation, the result is clear enough—no student is failed or conditioned. The average student receives a mark between 80 and 94; 67.8 per cent. of the class would be entitled to exemption from a final examination, while 18.3 per cent. of the students receive a mark between 95 and 100.

Chart 10 pictures the distribution of another course in the same college. The professor in charge, who believes in passing "handsomely," if at all, is in the habit of marking the papers strictly, and afterward deliberately raising the marks, so as to throw the entire class upward. The advancing of the marks is done by adding a small increment to

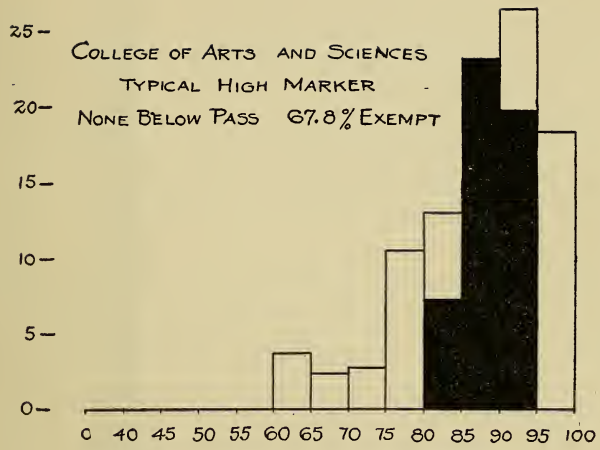


CHART 9. Distribution in Course B.

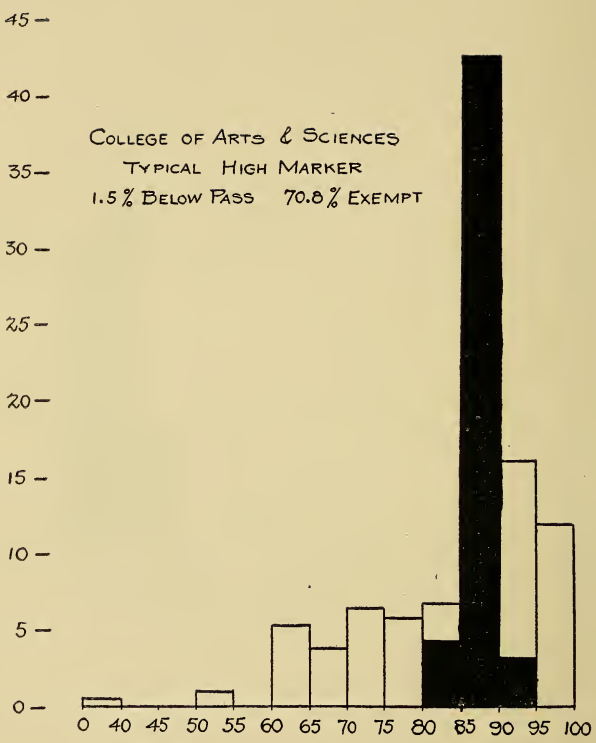


CHART 10. Distribution in Course C.

the original marks above 90, a larger increment to those between 80 and 89, a still larger one to the marks between 70 and 79, and so on. No marks are given between 55 and 59. A very small number of students are conditioned, and a still smaller number, practically negligible, are failed. A definite process of advancing medium-grade students to the exempt limit is also displayed. The result is clearly to produce an array in which the marks of 60 to 84 occur with almost the same frequency, in which the marks of 85 to 89 are given with disproportionate frequency, and in which altogether too many students are credited with an accomplishment of 90 or over. Exemption is granted to 70.8 per cent. of the class. Average students range between 80 and 94 in their marks.

Chart 11 shows an array derived from a course in the College of Mechanical Engineering. The primary explanation of the extremely favorable marks here is found in the reputation of the course as a 'snap,' to use students' parlance. It is also possible that the conditions under which the work is done may tend less than ordinarily to check cheating on the part of dishonest students. Whatever factor is at work, the result is that no student is conditioned or failed, that all average students would be entitled to exemption on the 85 basis, and that *93.6 per cent. of the students obtain a mark of 80 or over!* The remedy apparently is to condense the work into shorter compass or to incorporate the material, since it is so simple, in various other courses. The nature of the subject-matter suggests, at least, that it might well be included in the work given to the same students by another department.

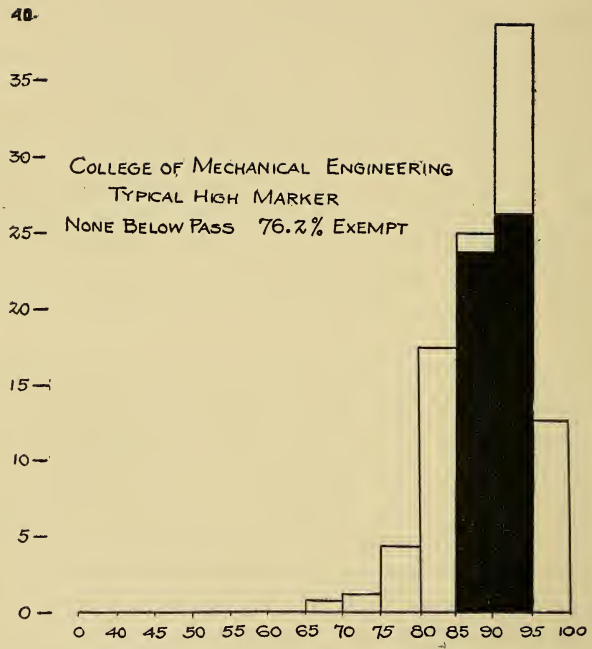


CHART 11. DISTRIBUTION IN COURSE D.

The most extreme case of high marking which came within the scope of this investigation is Course E, shown in Chart 12. The course is one in the College of Arts and Sciences. The work consists of lectures, outside readings and textbook study. The outside reading is not specifically prescribed, but must simply cover a given amount of time. In many cases the students read books that chance to interest them, but which do not bear directly upon the subject-matter of the course (though dealing, to be sure, with the general field of which the course forms a part). The lectures practically give everything found in the textbook, so that the latter might really be dispensed with by the student without affecting his standing. In addition, the professor in charge undoubtedly overestimates the accomplishments of the students, or, what amounts to the same thing in the end, sets a low standard of performance. The result is that no one is ever failed or conditioned, that 65 is the lowest mark given, that 78 per cent. would be exempt from the final examination, and that students of average ability are sure of a mark between 85 and 94.[1]

When we state that these four courses (B to E) are not selected, small-sized advanced classes, but large groups of undergraduates in elementary work, we can find no sufficient justification for the characteristics so strikingly exhibited in their graphs. Imagine groups of undergraduates, assembled at random from the student-body as they are in these

[1]The tendency to mark high is inherent in human nature. Dr. Ruffner says: "A temporizing professor who loves popularity, and desires, like the old man in the fable, to please everybody, is sure to be guilty of this fault, and, like many a politician, to sacrifice permanent good for temporary favor."

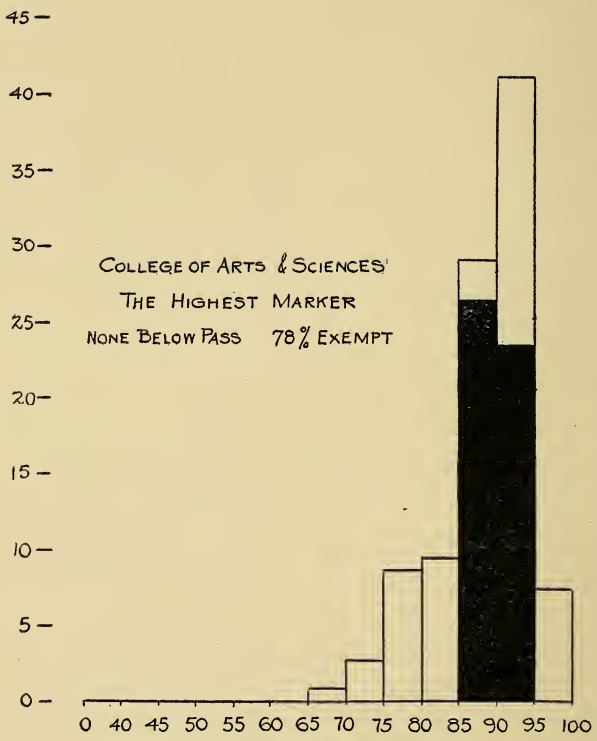


CHART 12. DISTRIBUTION IN COURSE E.

courses, who, year after year, display such extraordinary accomplishment! The marks charted in these four courses represent, collectively, 909 students, of which one unfortunate failed and two were conditioned. If all the courses of the university were patterned after these four, and if 85 were recognized, as it is now, as a mark of merit, then three out of four students would be entitled to election to the several honorary societies that seek for students of merit for enrollment in their organization.

3. Typical distributions of 'low markers.'

A selected group of 'low markers' is displayed in Table III, Courses F to M (Charts 13 to 19), with the exception of Chart 14, which is placed here for the sake of comparison.

Chart 13, which is perhaps the most extreme instance of low marking in our material, is a certain course in the College of Arts and Sciences which is

TABLE III.

EFFECT OF ZEAL AND DISTRIBUTIONS OF LOW MARKERS.

Showing the Per Cent. of Students in the Several Groups.

Course.	No. of Marks.	0-39.	40-44.	45-49.	50-54.	55-59.	60-64.	65-69.	70-74.
F	266	1.1	...	.7	5.3	...	27.0	32.0	12.4
G	353	...	...	1.7	7.3	.3	11.9	12.3	17.8
H	226	3.1	1.8	.9	4.4	.4	26.1	16.3	13.3
I	235	2.1	...	.9	1.7	4.2	23.4	14.5	19.2
J	234	2.1	.9	2.1	3.8	.4	21.8	16.7	19.1
K	317	1.2	2.5	2.5	5.7	5.7	18.3	9.5	12.0
L	208	.5	.5	1.9	3.8	1.9	18.7	10.2	16.4
M	273	...	...	1.5	2.9	6.9	17.6	17.9	22.7

Course.	75-79.	80-84.	85-89.	90-94.	95-100.	Exempt.
F	16.5	3.5	1.5	...	...	1.5
G	19.8	15.0	11.9	2.0	...	13.9
H	12.9	9.3	7.1	2.6	1.8	11.5
I	13.6	7.2	12.3	.9	...	13.2
J	8.2	6.1	14.1	4.7	...	18.8
K	9.8	6.6	16.7	6.6	2.9	26.2
L	14.9	9.6	12.1	8.1	1.4	21.6
M	13.6	12.9	2.9	1.1	...	4.0

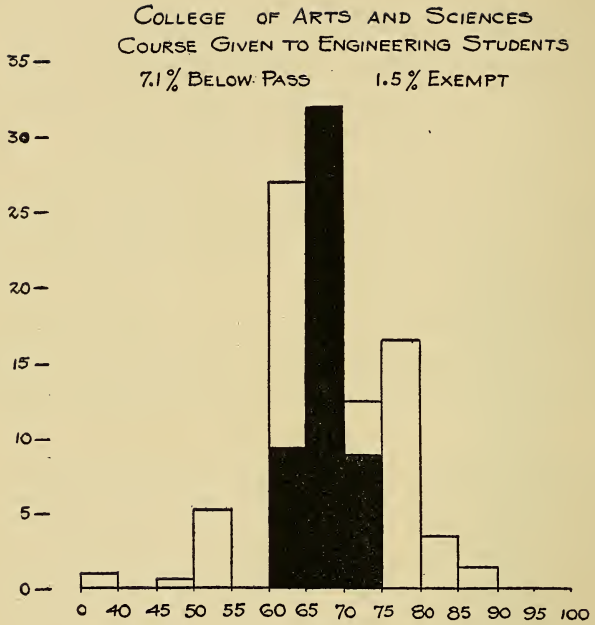


CHART 13. Distribution in Course F.

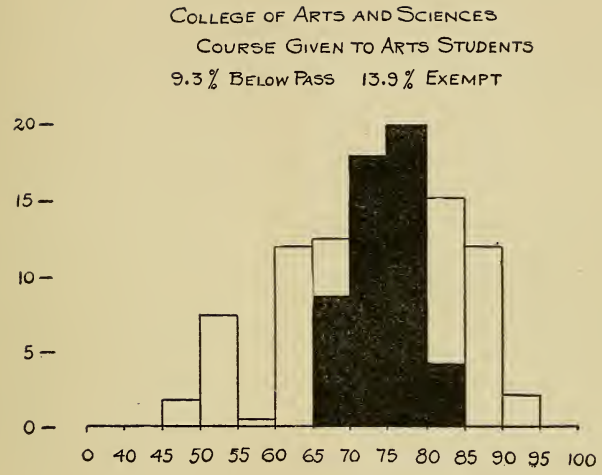


CHART 14. Distribution in Course G.

(For comparison with Chart 13, Course F, on the same subject.)

prescribed for students in engineering. For comparison with Chart 13 we introduce here Chart 14, which does not happen to belong to the low-marker group, but which is a course on the same general topic, in the same department, though given by another member of the instructing staff of the department and primarily to arts students. This latter course is more theoretical; the former, that given to engineers, is more practical and adapted to the immediate problems of the man in business. It is generally conceded that the former course is the easier, but the distribution of marks shows much poorer accomplishment in it. This situation is interesting enough to demand a moment's attention. The question arises: Why is the distribution of Chart 13 so different from that of Chart 14? The instructors who assigned these marks have been accustomed to work together in other courses, and their standards, so far as can be judged, are not dissimilar. Again, it is impossible to argue that the engineering students are, as a group, inferior in ability to the arts students. There is left, then, the factor of zeal or training. Evidence furnished by the testimony of numerous students and corroborated by the opinions of the instructors themselves, makes it clear that the engineering students, as a group, look upon this prescribed course as a 'necessary evil.' They take but a half-hearted interest in it, and, for the most part, strive merely to 'get through.' The result, as Chart 13 shows, is that no student receives a mark above 89, that only 1.5 per cent. are given marks of 85 or over, that the most frequent mark lies between 65 and 69, and that the middle half of the class fall between 60 and 74; in

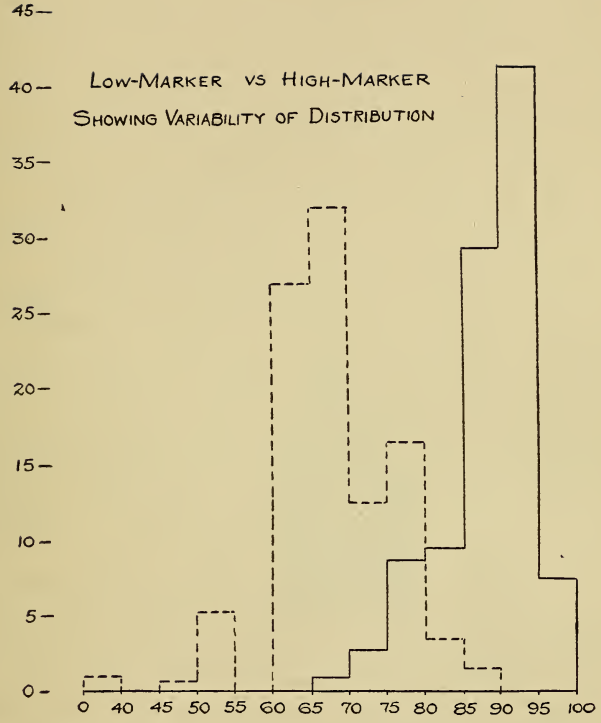


CHART 15. A Combination on the Same Base Line of Charts 12 and 13.

other words, fall, as a group, entirely below the average of the University.

Another point of interest: in both Chart 13 and Chart 14 there is displayed a decided aversion to giving a mark between 55 and 59; it is given but once in 619 cases. The examiner states that, in the subject in question, it is difficult to grade a student much closer than 5 points; and that it is impolitic to invite an argument over a mark between 55 and 59. This is further evidence of the impossibility of living up to the implications of the percentile system.

We pause here, before passing to other instances of low marking, to call the reader's attention to Chart 15, which, because it combines upon the same base-line the distribution of Chart 12 (high marker) and Chart 13 (low marker) will serve to picture graphically the inequalities of the regulation marking system in actual practice. This chart preaches its own sermon, so that further comment is unnecessary.

The marks displayed in Chart 16 are derived from a course in a department which has the general reputation of being the hardest marking department in the College of Arts and Sciences. Attention may be called to the mode at 60-64, to the progressive diminution of marks from 60 upwards, and to the fact that 10.6 per cent. of the students are conditioned or failed. A student of average ability cannot secure a mark above 79, while he may barely escape being conditioned.

Charts 17 to 19 (and likewise Course L, not shown graphically) are all examples of relatively low marking, combined with an interesting tendency to distribute the marks so that three modes appear.

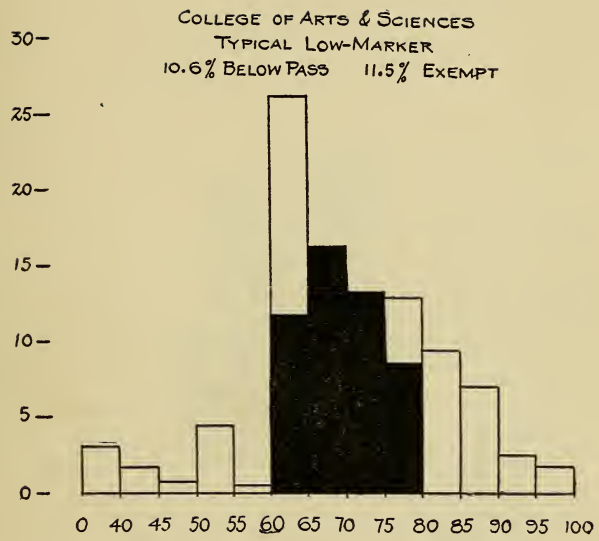


CHART 16. DISTRIBUTION IN COURSE H.

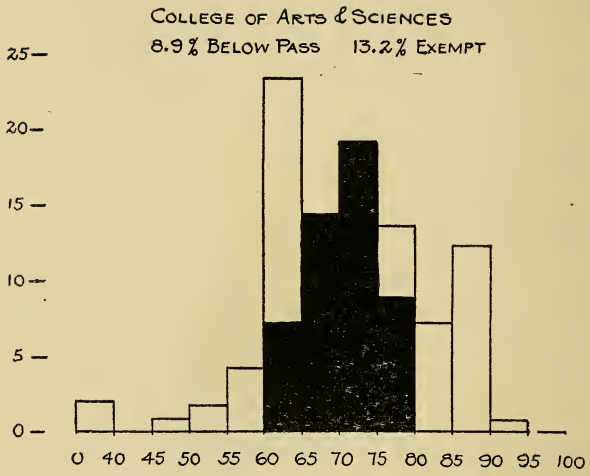


CHART 17. Distribution in Course I.

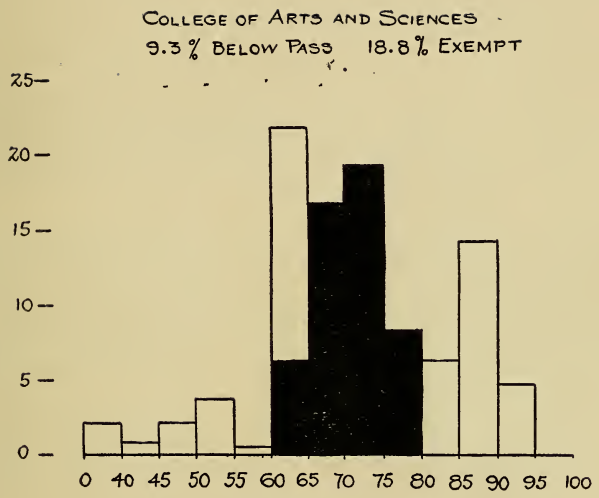


CHART 18. Distribution in Course J.
(Note the similarity to Chart 17, another course by the same instructor.)

In all of them the primary mode is at the 60-64 range; with one exception, the secondary mode is at the 70-74 range, and the tertiary mode at the 85-89 range. In other words, the commonest mark assigned is that which just permits the student to pass. The next most common marks are those which assign the students to a rank close to the average performance of the whole student body, while the next most frequent mark is that which entitles the student to exemption. Since more students barely pass than reach the average mark, the curves are all skewed to the left.

Charts 17 and 18 are of special interest, because they represent two different courses in the College of Arts and Sciences that are given by the same professor. Their similarity is striking, and is a good example of the influence of the personal equation in the distribution of marks.

In Chart 19 the distribution of Course K represents what may be termed a 'staff mark.' The course in question, given in the College of Arts and Sciences as a prescribed course to students in engineering, is divided into a large number of sections, and taught by a corps of teachers. The subject-matter of the course and the final examination are the same for all sections, but the marking of each section is in charge of its own teacher. All these marks are assembled in the one distribution. The curves show modes at the 60-64, 85-89 and 70-74 ranges; the average students' range is evidently too large (60 to 89); despite the mode at the 60-64 range, 26.2 per cent. of the class attain a mark of 85 or over. On the other hand, 17.6 per cent. of the students are conditioned

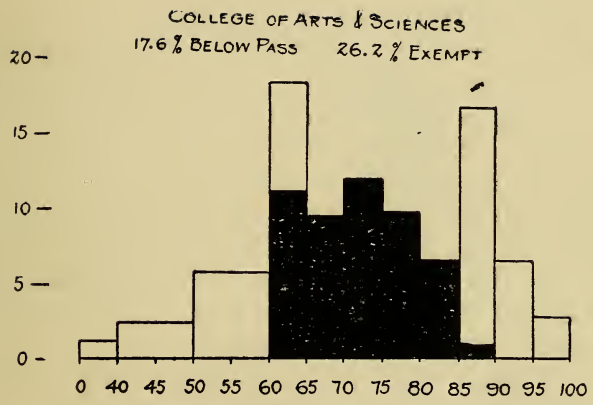


CHART 19. DISTRIBUTION IN COURSE K.

or failed. In short, the curve of distribution is too 'flat.'

Course L[1] shows practically the same shape of distribution as Course K. It has the three modes at the same ranges as the last chart, and has 21.6 per cent. of the class receiving exempt marks. The average student falls in the range from 60 to 84, and 8.6 per cent. of the students receive either 'condition' or 'failure.'

The distribution of Course M[1] is less decidedly that of a severely marked group. The mode falls at 70-74 rather than at 60-64, and the distribution falls away progressively on either side of this mode, as it should. However, the percentage entitled to exemption (4 per cent.) is quite small, and no member of the class gets 95 or over. This is a course in which the lectures are given by a series of different speakers, and the examination papers are marked under the supervision of a single man, who has general charge of the work of the course.

4. Peculiarities of distribution in other courses. The remaining distributions, in which the marking is neither decidedly low nor decidedly high, have been divided into two groups, the first comprising unimodal, the second multimodal distributions.

(*a*) The unimodal distributions, Courses N to T (see Table IV), have the merit of conforming, at least to some extent, to the theoretical distribution, according to which mediocrity or average accomplishment is the most frequent type. In the first four of these courses (only one of which, Course P, is here shown graphically), the range of marks as-

[1]Not here reproduced graphically. See Table III for details.

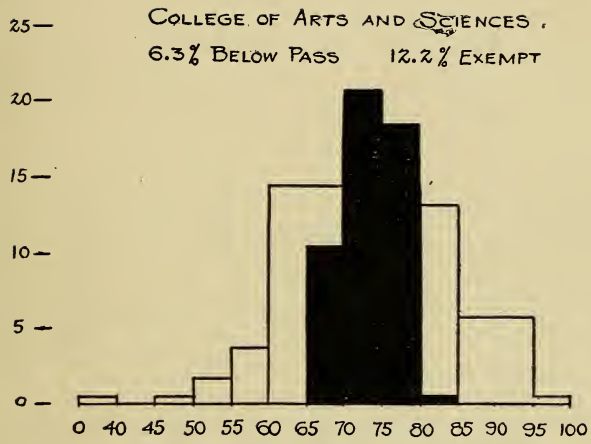


CHART 20. Distribution in Course P.

TABLE IV.

UNIMODAL DISTRIBUTIONS.

Showing the Per Cent. of Students in the Several Groups.

Course.	No. of Marks.	0-39.	40-44.	45-49.	50-54.	55-59.	60-64.	65-69.	70-74.
N.....	211	.9	.5	...	2.4	3.3	8.1	12.8	18.5
O.....	305	.7	.3	.7	1.6	1.9	13.1	14.8	16.4
P.....	221	.4	...	.4	1.8	3.7	14.5	14.5	20.8
Q.....	328	...	.3	1.2	3.9	4.6	9.5	17.1	18.0
R.....	254	.8	.8	1.9	.8	6.0	8.7	17.3	23.2
S.....	254	.4	...	...	1.5	.8	6.3	9.5	20.1
T.....	112	...	...	...	.9	...	...	2.7	9.8

Course.	75-79.	80-84.	85-89.	90-94.	95-100.	Exempt.
N.....	27.5	17.5	7.1	1.4	...	8.5
O.....	18.4	18.0	9.5	3.6	1.0	14.1
P.....	18.5	13.2	5.9	5.9	.4	12.2
Q.....	20.1	12.2	11.6	1.5	...	13.1
R.....	28.8	7.9	3.5	...	...	3.5
S.....	27.2	13.4	11.0	8.3	1.5	20.8
T.....	11.6	38.4	29.5	4.4	2.7	36.6

signed to students of average accomplishment is the same, viz., 65 to 84. In Course R the range is contracted to 65 to 79; in Course S advanced to 70-84, and in Course T to 80-89. In Courses R and S there is a slight irregularity at the 55-59 range, but the disturbance is too slight to disbar the distributions from the unimodal group. It is perhaps no accident that, with one exception (Course T), these unimodal distributions are derived from courses in pure or applied science.

(*b*) In our second group, the multimodal curves, are included Courses U to DD. (See Table V). In details they vary considerably. We shall pass over

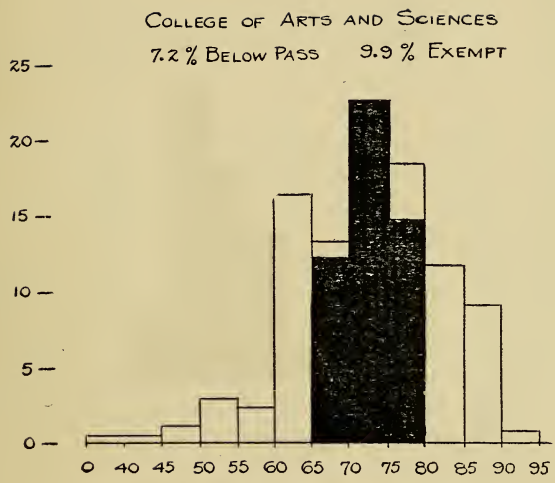


CHART 21. Distribution in Course U.

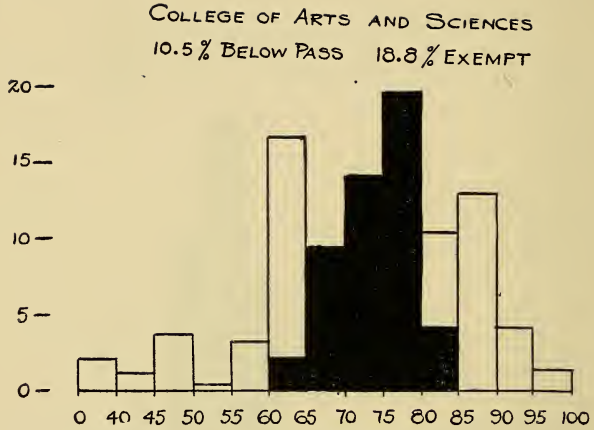


CHART 22. Distribution in Course V.

them rapidly, calling attention to features of interest in some of them.

TABLE V.

MULTIMODAL DISTRIBUTIONS.

Showing the Per Cent. of Students in the Several Groups.

Course.	No. of Marks.	0-39.	40-44.	45-49.	50-54.	55-59.	60-64.	65-69.	70-74.
U	262	.4	.4	1.1	3.0	2.3	16.5	13.4	22.8
V	191	2.1	1.1	3.7	.5	3.1	16.7	9.5	14.3
W	207	.5	...	1.4	2.9	.5	15.9	16.5	18.4
X	225	...	...	...	1.8	4.9	17.8	11.1	10.2
Y	251	...	.8	.4	5.2	3.2	14.7	13.7	18.6
Z	195	...	.5	.5	...	.5	1.5	1.5	9.8
AA	232	...	...	...	1.3	...	8.3	10.3	20.6
BB	215	...	...	1.4	5.5	2.8	15.7	19.6	16.3
CC	228	.4	.4	1.8	.4	2.2	3.1	7.9	18.0
DD	198	...	1.5	1.0	4.1	.5	10.6	13.2	11.6

Course.	75-79.	80-84.	85-89.	90-94.	95-100.	Exempt.
U	18.4	11.8	9.1	.8	...	9.9
V	19.8	10.4	13.1	4.2	1.5	18.8
W	17.4	4.3	19.3	2.9	...	22.2
X	15.1	7.1	18.6	11.6	1.8	32.0
Y	12.7	9.6	14.3	6.4	.4	21.1
Z	30.3	22.1	30.8	2.0	.5	33.3
AA	12.9	28.1	14.6	3.9	...	18.5
BB	18.7	11.2	7.0	1.0	...	8.8
CC	17.1	16.3	19.2	11.0	2.2	32.4
DD	15.1	13.2	17.2	7.5	4.5	29.2

Chart 21 is the array for a course in the College of Arts and Sciences, whose professor is generally reputed to be a fair marker. His curve of distribution varies, however, from the pattern we have recommended; first, in that it shows a general tendency to fall short in the frequencies in the upper end of the scale (no mark above 94, only 9.9 per cent. exempt, and no average student gaining a mark above 79); second, in that it shows a disproportion-

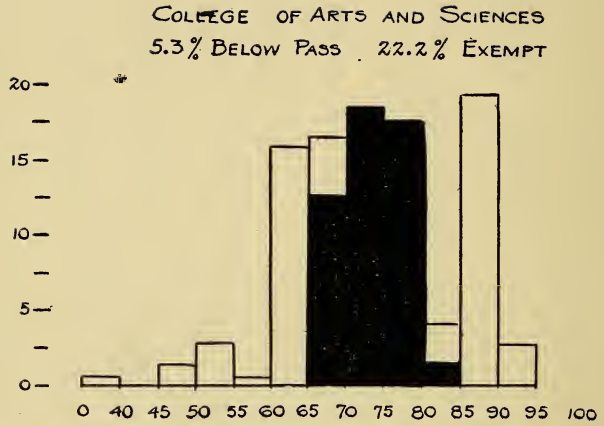


CHART 23. Distribution in Course W.

ate tendency to give marks between 60 and 64. Some of the marks in this range should have fallen in the 55-59 range, and some probably in the 65-69 range.

Chart 22 resembles Chart 19 in being a composite of marks for several sections of the same course, the present distribution being that of a language course in the College of Arts and Sciences, which is run in four sections. The two distributions have a certain amount of similarity, notably in the low frequencies assigned to the middle range of accomplishment. Medium-grade students range in marks from 60 to 84. The present curve has three modes, at 60-64, 75-79 and 85-89, respectively.

The peculiarity of Chart 23 (Course W), a science course, is in the curious 'hole' at the 80-84 section of the scale. To counterbalance this failing there appears a second curiosity—the range 85-89 forms the primary mode of the distribution. Again, despite the fact that 22.2 per cent. reach the grade of 85 or over, no one exceeds 94. Finally, the frequency of the marks between 60 and 79 is virtually constant for each section of five points.

Another language course is shown graphically in Chart 24 (Course X), which displays considerable irregularity and deviation from the distribution to be expected. Exemption is rather freely accorded (32 per cent.), and average students are spread over a range from 65 to 89. This failure to perceive the homogeneous character of this group of medium worth is the fundamental defect of the curve. A fondness for 60-64 is another evident characteristic, due, evidently, to pushing over the 60 mark those who should be conditioned, only 15 out of 225 are conditioned, while no one fails.

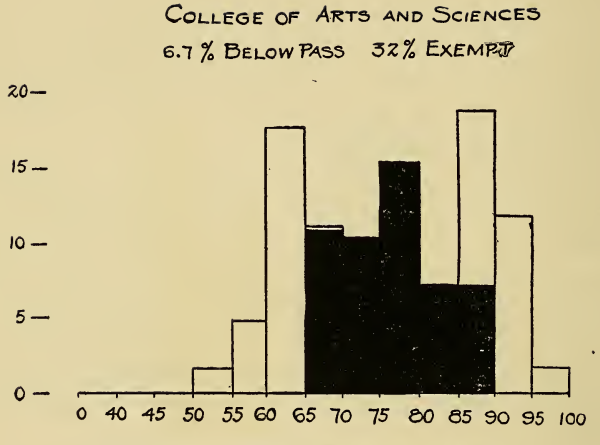


CHART 24. Distribution in Course X.

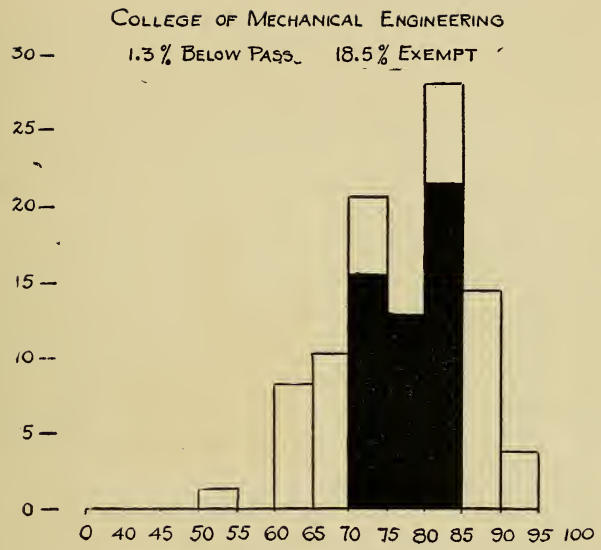


CHART 25. Distribution in Course AA.

A somewhat similar criticism could be passed upon Course Y (not reproduced graphically), a course in the College of Arts and Sciences, in which too many 80-to-84-grade students are pushed forward to exemption, and in which the 60-64 range is also too frequently used.

A member of the faculty who nearly deserves to be grouped in our second division, the high-markers, is responsible for Course Z (not reproduced, save in Table V). Exemption is gained by one-third of the class; only 1.5 per cent. of the students are failed or conditioned, while the average student never falls below 75, but may, indeed, win a mark of 89.

The course in the College of Mechanical Engineering (Course AA), displayed in Chart 25, is not one that should cause worry on the part of the student. While no one gets above 94, yet no one fails, and but 1.3 per cent. of the students are conditioned. The average student ranks between 70 and 84. Here the tendency is apparently to avoid the range 75-79 in favor of the range 80-84.

Course BB (see Table V) is from one of the large introductory courses in pure science. In general, the distribution inclines towards the lower marks, so that only 8.8 per cent. would reach exemption; no one exceeds 94, and 65-69 is the mode.

Another large elementary course in science is displayed in Chart 26 (Course CC). Save for a too high number of those exempted (32.4 per cent.), the distribution is one of the best found in our data.

In Chart 27, from the College of Arts and Sciences, the examiner has avoided the 55-59 range. The proportion exempted (29.2 per cent.) is too high, but the general form of the distribution is

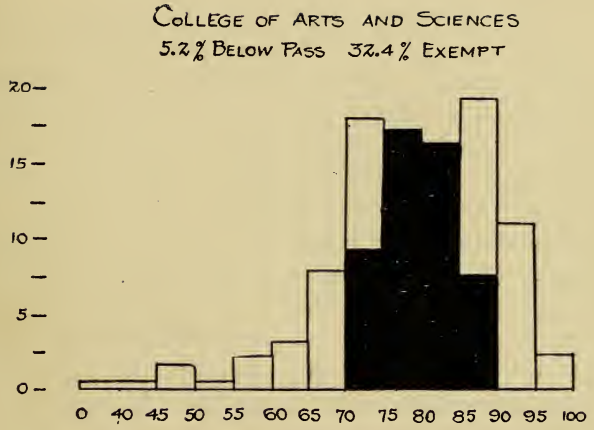


CHART 26. Distribution in Course CC.

otherwise not bad. This chart may be profitably compared with Chart 12 (p. 48), as the two come from the same department, apply to the same students, but are given by two different professors. The difference in the standards of accomplishment held by these two members of the faculty yields further evidence, if such be needed, of the inequalities that prevail in the marking system at present in force at Cornell.

5. The marking system of the College of Law. In the College of Law there prevails a system of marking that is radically different from the percentile system that we have just been discussing at length. The Law School system embraces six distinct marks, but, unfortunately, these six marks do not conform, either in intention or in practice, with the restricted-unit systems that were discussed in Chapter II. Nor does it appear that any effort has been made by those who introduced this system to relate it definitely to the percentile system which it replaced, and which still prevails elsewhere in the University. The translation into the regular University percentile system of the six symbols—*EE, E, G, P, P-60* and *Cond.*—which are in use in the College of Law, has been arranged in the charts which follow, in accordance with the statements of their values, as furnished by the Dean of that College.[1]

Law School Marks.....	*Cond.*	*P-60*	*P*	*G*	*E*	*EE*
Approximate Equivalents...	Below 60	58-63	60-74	75-89	90-98	99-100

[1]It appears that a mark (in the numerical system) of 60 might be either *P-60* or *P* in the College of Law. It is explained that the papers are marked in numerical terms and then translated into the six symbols. If a final paper warranted 62, the instructor would report the paper as *P* if the class work was acceptable, but as *P-60* if both class work and final examination were inferior.

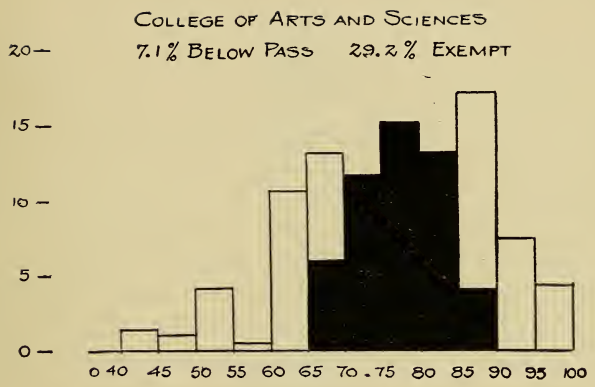


CHART 27. Distribution in Course DD.

(Compare with Chart 12 by another instructor in the same department.)

The data from the College of Law, Table VI and Courses EE, FF and GG, are derived from one first-year and two second-year courses. In examining

TABLE VI.

DISTRIBUTION OF MARKS IN THREE COURSES OF THE COLLEGE OF LAW.
Showing the Per Cent. of Students in the Six Groups.

Course.	No. of Marks.	Cond.	P-60.	P.	G.	E.	EE.	Ex-empt.
EE.....	251	20.3	17.2	33.6	21.8	7.1	...	7.1
FF.....	238	19.3	13.1	37.0	16.4	13.0	1.2	14.2
GG.....	222	12.1	16.3	36.5	27.9	6.3	.9	7.2

these charts the reader must remember that height of column, not area, is significant. For instance, the large areas on the left, which represent simply "below 60," are here plotted on the same abscissas used in constructing the graphs for other courses in the University. On the different charts *P-60* is represented by the range 60-64, while *EE* is represented by the range 95-100 (not strictly according to the numerical equivalence just quoted). Since no division is made at 85, it is impossible to compute the frequency of exemption on the same basis as for the other colleges. This frequency has been calculated, however, as if it included the marks *E* and *EE;* as if, in other words, it included marks of 90 and over. Naturally, this frequency is small—7.1 to 14.2 per cent.

We may begin with the consideration of Chart 28, the first-year course. The chief feature is the large number of students conditioned (20.3 per cent.), and the small number reaching the two upper grades (7.1 per cent. get *E,* none gets *EE*). What is the explanation of this extraordinary situation? From the evidence at our command it appears that the professor in charge is a severe marker, who be-

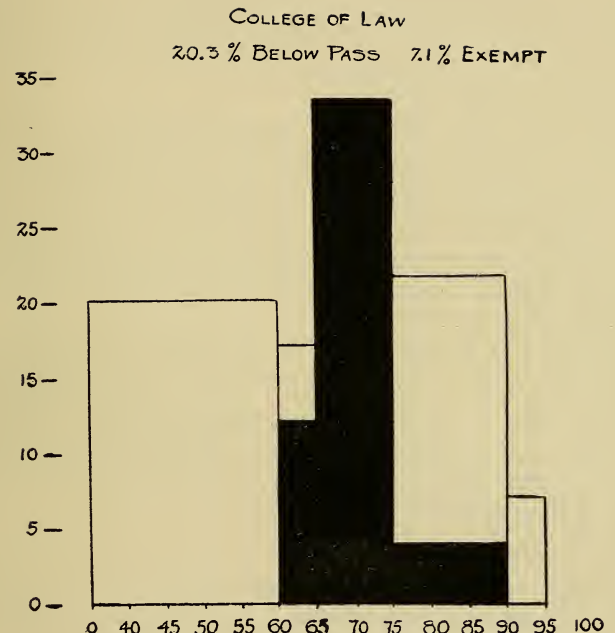


CHART 28. Distribution in Course EE.

lieves in conditioning regularly a certain percentage of his class; that the examination is searching; and, finally, that the nature of the work is different from what the beginner has encountered elsewhere in his career, whether in high school or college. The question may at least be raised whether some change ought not to be made in the conditions under which this course is given, so that not so many as one man in five would fail.

In Course FF (not shown graphically), a second-year law course, the conditions are practically identical, so far as conditioned students are concerned, and the curves elsewhere are closely similar, save that in this course the frequency of the two higher marks, *E* and *EE*, is somewhat increased at the expense of the mark *G*.

In our last graph, Chart 29, we show the distribution of another second-year law course (Course GG), whose examiner is reputed to be the highest marker in the College of Law. About 12.1 per cent. of the class is conditioned, but *P* remains the most frequent mark.

The trouble with these law curves is evident enough. They use a better number of symbols than other colleges in the University, but these symbols are, in our opinion, improperly rated. The most frequent mark is "Pass," which means inferior to the average, as we have seen in our discussion of the theoretical considerations underlying the use of a limited-division marking system. The mark *G* (good) should be changed to *M* (medium), or some other symbol of mediocrity, leaving two divisions above for superior and excellent students, and this mark *M* should be more frequent than *P*.

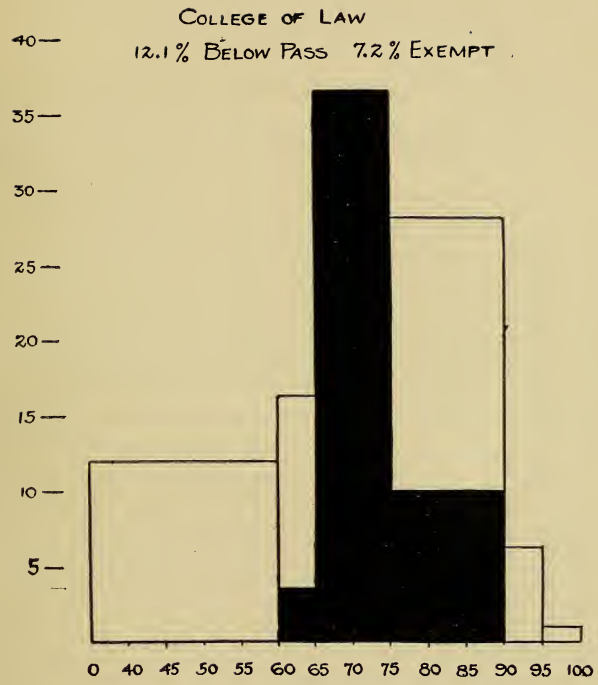


CHART 29. Distribution in Course GG.

CHAPTER V.

SUMMARY AND CONCLUSIONS.

Our investigation has led us to the following conclusions—some of them confessedly theoretical and deductive, others incontestable inductions from carefully compiled data.

1. The marking system of any institution of learning plays so important a rôle in the activities of the institution that its machinery, its significance and particularly its reliability is a matter that deserves and demands patient and impartial study.

2. Marks may be based upon performance, upon ability, or upon accomplishment. The last named is, save under unusual circumstances, the quality on which the marks should be based.

3. It is highly probable that ability, whether in high school or in college, is distributed in the form of the probability curve. It is at least possible, and we think it very probable that accomplishment, however, is distributed, under conditions commonly prevailing in school and college, in the form of a curve skewed toward the upper range.

4. The number of symbols proposed for recording degrees of accomplishment ranges from two to one hundred. Every theoretical consideration and many practical considerations favor a five-division system, based in essence upon five qualities of ac-

complishment, viz., excellent, superior, medium, inferior and very poor (failure).

5. A curve compounded from more than 20,000 marks shows that at Cornell University the 'pattern' distribution is that of a curve skewed toward the upper range, with a mode at 75-79, and the average at approximately 75 (60 being the pass-mark). The frequency of deviations above and below the mode decreases regularly on either side, save for a disturbance at the 60-point. This disturbance is caused partially by an effort on the part of some students to do just enough work to pass, but still more by a strong tendency of examiners to advance marks lying between 55 and 59 to 60 or over.

6. The data obtained for 31 individual courses (7430 marks) shows that the marks of members of the instructing staff are strongly affected by a personal equation—so much that typical distributions taken from high markers and from low markers show no similarity whatsoever.

a. The percentage of students obtaining 85 or over (a range which, in many classes, entitles the student to exemption from final examination, and which, by assumption, indicates a quality of work superior to that of the medium student) falls to 1.5 per cent. in one class, and rises to 78 per cent. in another class in the University.

b. Students of medium accomplishment (who by definition are relatively like one another in merit) are by some examiners rated between 85 and 94, but by other examiners 60 to 74. Again, these students are by some instructors spread over a range of 30 points, by others limited to a range of 10 points.

c. The marks of the same students, continuing the

same subject, show a different form of distribution when the instructor is changed.

d. Distributions which show radical divergencies in form and tendency may be obtained from the records of two teachers engaged in precisely the same work.

7. These and other variations in the assignment of marks need not always be laid at the door of the instructor. We have shown how the same subject, taught to different groups of students, *e. g.,* to arts students and to engineering students, may yield a differently formed curve of distribution.

8. The curves for individual courses are often multimodal. In other words, there are two or more ranges in the marks which occur with a frequency greater than that of the ranges on either side of them. Commonly, these modes are located at three points, viz., 60-64, 75-79 and 85-89. The first of these is due to the tendency indicated above (Conclusion 5): the second is the normal 'peak' of average accomplishment; the third is due to a tendency, analogous to the first, to increase the number of students who are exempt from final examination, *i. e.,* to advance marks from 80-84 to 85 or over.

9. There appears to be a tendency for marks in courses in pure science and applied science to conform more closely to the theoretically presumptive distribution than do marks in other courses. But this generalization is insecure because, after all, we have charted in detail only 34 out of the several hundred courses offered in the University.

10. The marking system employed in the College of Law has the merit of using a restricted number of symbols, but it does not conform to the theoretical

curve of distribution, nor was it designed with the proper theoretical considerations (discussed in Chapter II).

11. The marking system used by most faculty members for recording the work of graduate students (two divisions, satisfactory and not satisfactory,) is not to be recommended for use with undergraduates, at least under the conditions that now prevail.

12. *We recommend that every institution of learning, at least every high school and college, adopt a five-division marking system, based upon a distribution which should, in the long run, not deviate appreciably from the following: Excellent,* 3 *per cent.; superior,* 21 *per cent.; medium,* 45 *per cent.; inferior,* 19 *per cent; very poor,* 12 *per cent.* For purposes of administration the very poor group may be subdivided so that approximately 11 per cent. shall be conditioned, and 1 per cent. shall fail. This distribution conforms well with theoretical requirements, and coincides closely with the present practice of Cornell University, as shown by the tabulation of 20,348 marks, drawn from a period of three different years and from 163 courses. It is important to note that, by this proposed system of marking, the meaning of each mark is exactly defined, and in the only satisfactory way by which a mark can be defined, viz., in terms of the frequency with which it can be secured by students under actual working conditions.

13. Furthermore, as Meyer (9, p. 664) advocates, in order to ensure the working of the system the

distribution actually given should be tabulated at stated intervals, say biennially, and the distribution should be made public, so that every examiner shall know to what extent he conforms to the principles on which the system is based.

BIBLIOGRAPHY.

(1) CATTELL, J. M. Examinations, Grades and Credits. *Popular Science Monthly*, 66: 1905. p. 367.

(2) COLVIN, S. S. Marks and the Marking System as an Incentive to Study. *Education*, 32: 1912. May, p. 560.

(3) DEARBORN, W. F. School and University Grades. *Bulletin of University of Wisconsin*. 1910, No. 368.

(4) FOSTER, W. T. Administration of the College Curriculum. Boston: H. Mifflin Co. 1911. Chap. 13.

(5) HALL, W. S. A Guide to the Equitable Grading of Students. *School Science and Math.*, 6: June, 1906.

(6) HUEY, E. B. Retardation and the Mental Examination of Retarded Children. *Journal of Psycho-Asthenics*, 15: Sept. and Dec., 1910, p. 31.

(7) JUDD, C. H. On the Comparison of Grading Systems in High Schools and Colleges. *School Review*, 19: 1910, p. 460.

(8) MEYER, M. The Grading of Students. *Science, N. S.*, 28: 1908, p. 243.

(9) MEYER, M. Experiences with the Grading System of the University of Missouri. *Science, N. S.*, 33: 1911, p. 661.

(10) SARGENT, E. B. Education of Examiners. *Nature*, 70: 1904, p. 63.

(11) SMITH, A. G. A Rational College Marking System. *Journal of Educational Psychology*, 2: 1911, p. 383.

(12) STEELE, A. G. Training Treachers to Grade. *Pedagogical Seminary*, 18: 1911, p. 523.

(13) STEVENS, W. L. American Titles and Distinctions. *Popular Science Monthly*, 63: 1903, p. 310.

INDEX.

Accomplishment, how determined, 10; distribution of, 14 ff., 79.
Arts and Sciences, College of, 42 ff., 47, 49 ff., 65 ff.
Cattell, J. McK., 5, 14, 19, 27, 30, 31, 32.
Charts, general explanation of, 23 ff.
Colvin, S. S., 6.
Conclusions, 79 ff.
Cornell University, marking system in, 15 f., 23; distribution of marks 1902-03, 21 ff.; 1903-04, 21 ff.; 1910-11, 22 ff., 37 ff.
Dearborn, W. F., 5, 14, 19, 27, 31, 32.
Distribution of marks, theoretical, 11 ff., 79 ff.; ideal, 33, 82; unimodal, 60 ff.; multimodal, 62 ff., 81. See high-markers; low-markers.
Elimination, effect on distribution, 12 ff.
Foster, W. T., 5.
Hall, W. S., 5, 33 ff.
High-markers, 42 ff.
Judd, C. H., 13.
Law, College of, 72 ff., 81 f.
Low-markers, 49 ff.
Marking system, theory of, 9 ff.; two division, 16 f., 82; three division, 17 f.; four division, 18; five division, 18 f., 30 ff., 79 f., 82; Percentile, 6, 19 f.
Mechanical Engineering, College of, 45, 69 f.
Missouri, University of, 14, 20.
Meyer, M., 5, 6, 7, 13, 14, 19, 20, 27, 30, 31, 32, 82.

Native ability, distribution of, 11 ff.
Northwestern University Medical College, distribution in, 33 ff.
Performance, marks based on, 9 f., 79.
Probability curve, 11.
Ruffner, 47.
Science, courses in, 62, 70, 81.
Skewed curve, 12 f.; interpretation of, 31 ff.
Smith, A. G., 5, 10 f., 14.
Starch, D., 20.
Steele, A. G., 27.
Tables, I, 22; II, 38; III, 49; IV, 62; V, 65; VI, 74.
Units, best number of, 16 ff., 79 f.
Variations produced by change of instructors, 39 ff., 80 f.

SEP 17 1913

LIBRARY OF CONGRESS
0 029 448 208 8